

On the precision of third person perspective augmented reality for target designation tasks

Fei Liu^{1,2} · Stefan Seipel^{1,2}

Received: 11 December 2015 / Revised: 22 July 2016 / Accepted: 28 July 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The availability of powerful consumer-level smart devices and off-the-shelf software frameworks has tremendously popularized augmented reality (AR) applications. However, since the built-in cameras typically have rather limited field of view, it is usually preferable to position AR tools built upon these devices at a distance when large objects need to be tracked for augmentation. This arrangement makes it difficult or even impossible to physically interact with the augmented object. One solution is to adopt third person perspective (TPP) with which the smart device shows in real time the object to be interacted with, the AR information and the user herself, all captured by a remote camera. Through mental transformation between the user-centric coordinate space and the coordinate system of the remote camera, the user can directly interact with objects in the real world. To evaluate user performance under this cognitively demanding situation, we developed such an experimental TPP AR system and conducted experiments which required subjects to make markings on a whiteboard according to virtual marks displayed by the AR system. The same markings were also made manually with a ruler. We measured the precision of the markings as well as the time to accomplish the task. Our results show that although the AR approach was on average around half a centimeter less precise than the manual measurement, it was approximately three times as fast as the manual counterpart. Additionally, we also found that subjects could quickly adapt to the mental transformation between the two coordinate systems.

✉ Fei Liu
feiliu@hig.se
Stefan Seipel
ssl@hig.se

¹ Department of Industrial Development, IT and Land Management, University of Gävle, 80176 Gävle, Sweden

² Centre for Image Analysis, Uppsala University, 75105 Uppsala, Sweden

Keywords Augmented reality · Third person perspective · Target designation · Precision study · Experiment

1 Introduction

Augmented reality (AR) combines the real world with computer-generated information in real-time. The virtual information is registered with related real objects to provide users with better understanding of those objects. The potentials of this new paradigm of visualizing and interacting with the actual surroundings have been enticing a large body of research and practice with regard to its applications. In Azuma's influential survey paper [2], the application areas he reported covered only medicine, manufacture of complex machinery and military. After thirteen years, Van Krevelen and Poelman [32] published an updated survey in 2010 and their work revealed that through these years AR applications had not only become more sophisticated in the traditional fields but also spread widely into education, offices, personal information and entertainment. With the maturation of enabling technologies as well as the ensuing decrease of costs, barriers of adopting AR are being swept away and therefore we can expect to see more reliable AR applications reaching a wider audience.

The industries of Architecture, Engineering, Construction and Facility Management (AEC/FM) need to work with plenty of information during the process of a project. Although information digitalization and integration do improve the work efficiency and reduce the costs, one persisting bottleneck of performance in construction practices is the disconnection between the cyber information and the related physical world elements [35]. For instance, a site manager needs to mentally compare a digital building model against its real counterpart in order to determine if it is constructed as planned. This is a time-consuming and error-prone process, which is likely to cause further project delay, quality degradation, cost overrun as well as other harmful consequences. AR is a natural solution to overcome this bottleneck. In fact, Shin and Dunston [13] identified eight work tasks, namely, layout, excavation, positioning, inspection, coordination, supervision, commenting and strategizing, in the field of AEC which can potentially benefit from the application of AR. Researchers over the world have also been contributing numerous works to prove the suitability of AR in AEC/FM and they were thoroughly surveyed in [4, 7, 24].

Thanks to the advancement of microelectronics, it is very common for today's consumer mobile devices to boast high speed powerful processors, high resolution LCD screens and integration of various sensors such as mega-pixel cameras, WiFi network access, inertia-measurement unit (IMU), compass and GPS [10]. On the other hand, off-the-shelf AR software toolkits, for example, Metaio, Vuforia and ARtoolkit, provide state-of-the-art solutions for tracking, registration and rendering, which are essential to any AR applications. The accessibility of both hardware and software prerequisites has made handheld video see-through AR a strong candidate for AR-based solutions in the AEC/FM domain.

However, due to the limited field of view of built-in cameras [17], AR users usually have to move away from the augmented object in order to 1) capture enough features (whether natural or artificial) for a successful tracking and 2) view the virtual overlay as much as possible. This is particularly true for AR applications in the field of AEC/FM since they tend to deal with large-sized objects, e.g., walls and façades, and the distance to move away can be rather significant. This creates a problem when the users also want to physically interact with those objects at the same time, which commonly employed first person perspective (FPP) AR is unable to tackle. For instance, a facility maintenance fieldworker may want

to mark on the wall the locations of hidden pipes according to the virtual pipes displayed by a tablet-based AR tool or a thermographic inspector may try to delimit a large defected region on a façade through her handheld AR visualization tool. One solution to this problem is utilizing AR from the third person perspective (TPP). Considering the aforementioned example of maintenance fieldworker, a stationary camera can be placed far enough to capture the wall. The video stream is then sent to a mobile smart device held by the worker and augmented with virtual pipes on the device. Since the worker will be in the camera view as well, she can guide herself to perform desired interaction tasks with the help of the augmented video stream. One supplementary advantage of adopting TPP AR over the first person close-up AR is that the former grants users a broader context of the targets to be marked, which may result in reduced search time and therefore improved performance.

On the other hand, TPP AR requires mental transformations between the user-centric coordinate space and the coordinate system of the remote camera. The objective of our study is to evaluate user performance of this cognitively demanding AR approach with respect to a fundamental 2D target designation task. We envision that the study results would provide insights into the viability of applying TPP AR to potential tasks in building construction and maintenance. To this end, we implemented an experimental TPP AR system and designed an abstract-scenario experiment which required subjects to make markings on a whiteboard according to virtual marks displayed by the AR system. The precision and the time of marking were measured for this study. Details about the enabling system and the experiment are described in Section 3 and 4 respectively. We then present the results in Section 5 while Section 6 discusses the implications of the results, the limitations of the TPP AR system and concludes the study.

2 Related work

The first stand-alone handheld AR system built upon consumer-level devices was implemented by Wagner and Schmalstieg [34] in 2003. The system utilized a personal digital assistant (PDA) attached with a commercial camera as the client and they demonstrated the feasibility and the effectiveness of the system. Since then, many handheld AR systems based on more and more powerful mobile smart devices have been developed for various applications. For the general public, applications for tourism and navigation have received much attention. Keil et al. [16] presented an AR-based tourism guide for historical sites. Handheld AR applications as museum guide have been reported in [9] and [6]. As for personal navigation, Mulloni et al. [21] proposed a system consisting of sparse *info points* which are connected with activity-based instructions. The indoor evacuation system described in [1] computes an optimal exit path based on user's current position, floor layout and personalized pedometry and then displays it on a mobile device in the form of AR. The work of Dünser et al. [11] evaluated the usefulness of handheld AR browsers for guiding people to points of interest with respect to digital map-based interfaces. Pedestrian navigation with AR, voice and digital map were compared in [25].

Handheld AR has also been actively explored in the community of AEC/FM in recent research. Bae et al. [3] developed a vision-based mobile AR system which allows fieldworkers to access co-registered building information modeling (BIM) information on-site with high precision. The project Smart Vidente [27] involved a handheld AR prototype with centimeter-level registration accuracy for utility companies to perform daily surveying of geospatial objects such as underground infrastructure. An AR solution was described

in [23] to enable stakeholders to gain mobile access to BIM information on-site as well as making annotations for updating and sharing. InfoSPOT [14] is a mobile AR tool for enhancing the situation awareness of facility managers and [18] brought forward a mobile defect management AR application which facilitates construction workers and managers to detect dimension errors and omissions on the worksite.

There exists research concerning AR-based measurement tools as well. Experiments conducted by Kamat and El-Tawil [15] show that measuring interstory drift ratio of a building after a seismic event through AR techniques deserves attentions of field implementation. Shin and Dunston [12] compared the results of steel column inspection obtained from an AR prototype system and a total station. A survey on AR as user interface for measurement systems can be found in [10].

While AR systems, especially handheld AR, are typically designed in the first person perspective, TPP AR is not at all a new concept, for example, the world-in-miniature aid in [5]. Milgram and Colquhoun [19] introduced a continuum to represent different types of viewpoint centrality. Tönnis et al. gave the definitions of three key classes on the continuum: egocentric, egomotion and exocentric in their survey paper [31]. AR interfaces presenting the physical world and/or the virtual contents from multiple exocentric viewpoints have been explored in [28] and [30]. With this kind of interfaces, users can view and manipulate the virtual contents without physically being at those vantage points. Transitioning techniques between the egocentric and the exocentric views have also been reported [20, 29, 33]. The focus of all these studies is AR interaction techniques based on multiple viewpoints and the accompanying user experiments were performed to validate the concepts. Our work does not involve multiple viewpoints and hence no view transitioning technique is needed. Rather, our emphasis is user task performance associated with a single exocentric view. Therefore, this study is more in line with the one from Salamin et al. [26], although the user tasks investigated are different. In order to find out the benefits of both first and third person perspectives in virtual and augmented reality, Salamin et al. [26] designed and carried out experiments in which subjects needed to complete, among others, navigation and locomotion tasks with both perspectives. Another system with the similar design as ours was proposed by Côté and Trudel [8] except that they tracked both the orientation of the handheld device and the position of a panoramic camera which provided the third person view. According to the authors, the use of a panoramic camera offered easier tracking and more accurate registration between the real world and the virtual models. On the other hand, tracking the orientation of the handheld device, which is not a part of our implementation, allowed the users to view the virtual model from different angles without moving the panoramic (remote) camera. Their work, however, lacks experiments to validate user performance with the system.

3 TPP AR system

As stated in the introduction section, the general idea of such TPP AR systems is to capture both the user and the whole object to be augmented with a remote camera and then present the captured video frames composited with virtual information to the user via a mobile smart device (see Fig. 1). In our study, for improved performance, we chose to offload the tasks of rendering the virtual information and integrating it with the video frames of the real scene from the smart device onto a laptop computer, Acer Aspire V5-573G. The laptop is equipped with Intel Core i7 1.8GHz CPU, 8GB RAM, Nvidia GeForce GT 750M graphics card and 15.6" HD display with the optimal resolution of 1366×768 . A Logitech

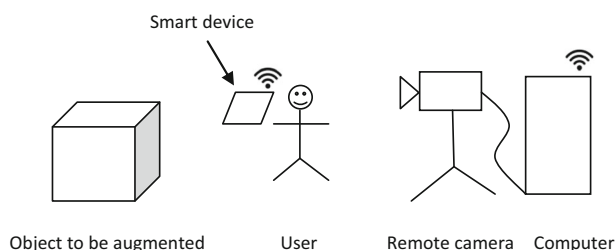


Fig. 1 Illustration of the main idea of the TPP AR system used in this study

C270 webcam (supporting up to 1280×720 video resolution) was attached to it for scene capturing since the built-in camera had rather poor quality. The augmented video stream was then sent to a smartphone, Sony Xperia Z3 Compact, through a WiFi local network. The smartphone runs Android 4.4 operating system with a quad-core 2.5 GHz CPU, 2GB RAM and 4.6" display at the resolution of 720×1280 . We implemented our AR application with Unity game engine (version 5.0) for its rapid, cross-platform development supports and Qualcomm Vuforia mobile vision platform (version 4.0) for AR-specific functionality. An overview of the process flow of our system is depicted in Fig. 2.

With our TPP approach, the user can stand within the vicinity of the physical object to interact with it while accessing the augmented view as if she was away from the object. Meanwhile, the user can also take advantage of her position relative to the object in the video as guides to perform interaction tasks such as designating 2D positions indicated by the AR system, which was chosen to study the effects TPP AR has on user performance. To support this task, a rectangular region was prepared on a whiteboard as the area to be augmented with virtual position symbols (the area will be called workarea henceforth). Vuforia offers marker-based tracking and registration as well as removal of lens distortion based on an initial lens calibration procedure. The markers used by Vuforia are called frame markers. Each frame marker has a unique identifier which is encoded through the black-and-white pattern on the edges of the marker and there are 512 markers available in total. The first four markers (Fig. 3) were used in this study and we printed them out with the size of 16×16 (cm). Such a big size ensured they were still discernible at a distance. Marker users can add any human-readable information, e.g., text or pictures, at the center of the markers to tell them from each other. In this study, as shown in Fig. 3, we printed the marker numbers in the center for this purpose. Figure 4a shows the arrangement of these four markers related to the workarea. Note that one of the four corners of a marker aligns precisely with a corner of the workarea. For example, the top right corner of Marker0 is also the bottom left corner of the workarea etc. These four markers are independent of each other and Vuforia is capable of correctly augmenting the workarea with any one of them

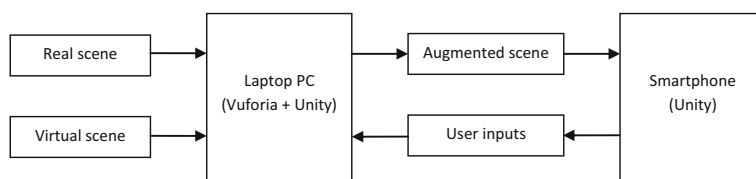
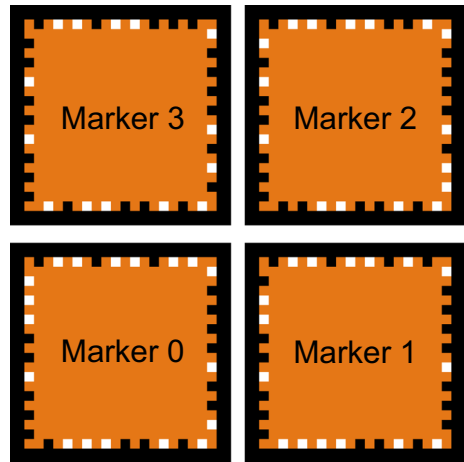


Fig. 2 Overview of the process flow of our system

Fig. 3 Frame markers used in this work



alone. The reason why four markers were employed is to make sure the system continues to function when users occlude some of them during the experiments.

Our AR application comprises two parts. The main part of it runs on the laptop computer, which contains a virtual counterpart of the workarea. It was modeled at true scale in Unity with Vuforia as its plug-in. As we can see in Fig. 4b, the four virtual markers were arranged in the same way as the real ones. Virtual objects for augmentation include a red cross to indicate the current target for marking, an optional rectangular grid with neighboring lines 10 cm away from each other in both horizontal and vertical directions and lastly four red squares, which indicate the marker that is being tracked by Vuforia at the moment. Once Vuforia completes rendering the virtual scene and integrating it into the live video stream, the application converts each augmented frame into a JPEG image and sends it to the smartphone for display. To maintain interactive co-registration of the video feed with the virtual objects in Vuforia, we had to downsample camera images to a quarter of their original size (i.e. 640×360 pixels). We achieved 30 frames per second by doing so, which required a network communication bandwidth of approximately 4 MB/second. Figure 5 shows a screenshot of the smartphone application, which is the second part of our AR application and here we can see the augmented video stream as well as a simple user interface (UI). The major UI elements are a text label on the top left and a pair of buttons at the bottom. The text label displays the current target number and its coordinate while the two buttons

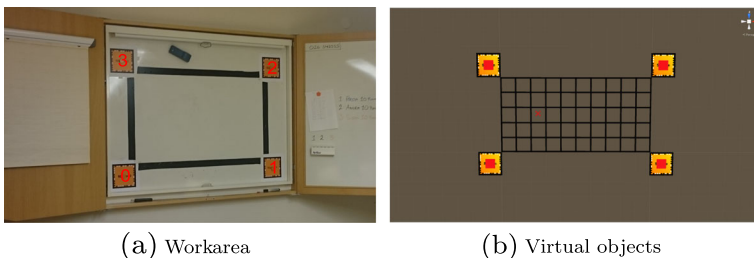


Fig. 4 The actual workarea with markers numbered and virtual objects for augmentation



Fig. 5 A screenshot of the smartphone application

allow users to navigate sequentially through a list of targets to be marked thus controlling the experiment flow.

4 User experiments

The goal of the experiment is to evaluate the usability of TPP AR for a generic facility maintenance task, namely the designation and the marking of targets on walls. Consequently, we recorded both positions marked by users and time they took to mark each target as quantitative measures from the experiment. Additionally, subjective opinions were also gathered as qualitative measures in the form of post-session questions and comments concerning learnability and subjects' feeling towards the TPP.

4.1 Subjects

In order to verify experimental results and to be able to generalize to a larger group of users, we managed to carry out the user experiment in two different environments with slightly dissimilar test populations: the experiment was first performed in the computer science department of our university with User Group 1 and one week later repeated in a technology consulting company which provides customers with electronic and software solutions with User Group 2. Subjects from Group 1 were mainly students and a few members of staff in the department and some of them had already had experience with developing and using AR applications. On the other hand, subjects from Group 2 were mostly software or hardware engineers (only two were administrative workers). Therefore, in general both groups of people had strong technology background, while Group 2 comprised entirely test persons with practical working experience. In total we recruited 22 subjects, 12 in Group 1 and 10 in Group 2.

4.2 Setup and procedure

Due to the differences between the two rooms as well as the sizes of whiteboards we had access to during the experiments, the dimension of workarea and the distance from the camera to it had to be varied in these two locations. The workarea in the university was 100.3 cm in width and 69.6 cm in height while the one in the company was 100 cm in width and

60 cm in height. In both locations, the center of the workarea was about 150 cm above the floor level, which ensured all subjects could reach any point within the workarea naturally. The camera was placed 220 cm away from the workarea in the university setting and at the distance of 320 cm in the company. Accordingly, the virtual counterpart of the workarea as represented in the Unity 3D environment was altered depending on the experiment locations.

The experiment was conducted according to the same procedure in each group. Every subject in a group performed three trials to carry out a target designation task. Within each trial, subjects were asked to mark a set of pre-determined targets on the workarea with a pen. The trials differed in the method of identifying the target positions within the workarea. In Trial 1, subjects were only given a ruler and a list of coordinates so they had to locate every target manually. The purpose for this trial was to mimic the traditional style of FM maintenance fieldwork and to provide a contrast to our AR approach. For the remaining two trials, the subjects had access to the AR system and it would sequentially augment the workarea with a red cross at a pre-determined position at a time to aid subjects in determining and marking these targets on the workarea. Figure 6 shows a subject performing Trial 2 or Trial 3 with the AR tool. The difference between these AR-based trials was that Trial 2 did not render the virtual grid, while Trial 3 had one. The grid, as mentioned in the previous section, had an interval of 10 cm between adjacent lines of the same orientation and therefore acted as a reference of the underlying coordinate system. Its use was intended to study if such an explicitly drawn coordinate system could help users mark the target more precisely. For an example view of the AR tool from a user's perspective in Trial 3, please refer back to Fig. 5. The trial order was the same for all subjects in both experiments but different sets of stimuli (i.e. target positions) were used for the three trials in order to reduce potential learning effects. Tables 1 and 2 list target positions we used in experiments conducted in

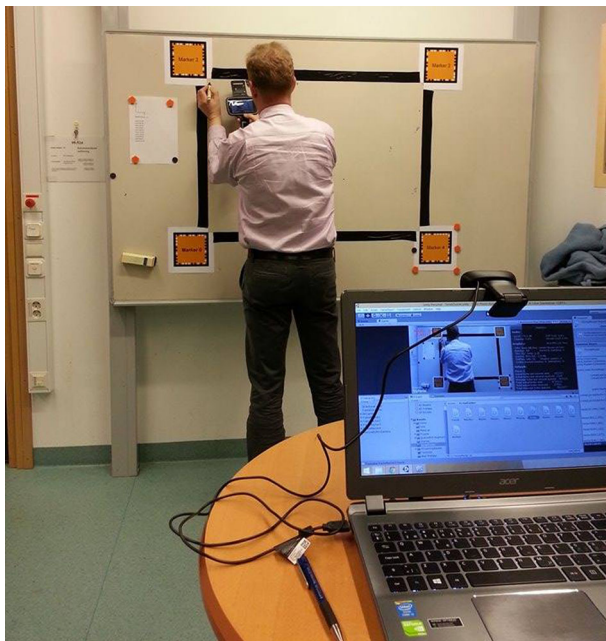


Fig. 6 A subject is guiding his pen to mark the designated position with the help of augmented video displayed on the smartphone in Trial 2 and 3

Table 1 Target coordinates (x,y) in cm for Group 1

	Target1	Target2	Target3	Target4	Target5	Target6	Target7	Target8	Target9	Target10	Target11
Trial 1	(52,20)	(85,38)	(39,15)	(40,67)	(96,18)	(21,46)	(14,62)	(97,47)	(26,4)	(16,37)	–
Trial 2	(45,37)	(63,25)	(98,57)	(2,69)	(36,46)	(27,4)	(71,26)	(69,67)	(39,28)	–	–
Trial 3	(84,11)	(42,48)	(59,39)	(30,63)	(82,37)	(54,20)	(48,61)	(16,30)	(13,55)	(87,25)	(88,4)

both locations. The coordinate system was defined as follows: the left bottom corner of the workarea was the origin with the positive x pointing towards the right while the positive y pointing upward.

Before each subject began, we introduced to the subject what the objective of the experiment was and what she should do in each trial. The subjects were instructed to locate the targets at their own pace, which means no external requirements for precision and completion time were imposed. Once the subjects started their experiments, we did not intervene in the process via, e.g., offering help with the designation. The subjects completed all three trials independently. The experimenters only handed the mobile phone to the subjects when they finished Trial 1 so that they could continue with Trial 2 and 3. Also, no special user-testing protocol was adopted for this study. During the manual designation trial, a stopwatch was used to record the marking time. When the subject began to locate a target, the experimenter started the stopwatch and when the subject thought she had successfully marked the target on the workarea, she would signal the experimenter to pause the stopwatch and the latter would then make a note of the time. This timing mechanism was implemented in the AR system utilizing the flow-control buttons (“previous point” and “next point” mentioned above) so no manual timing was needed in the second and the third trial. After completion of each trial, the experimenter measured the coordinates of all pen marks on the workarea with the help of a ruler and a cross laser pointer (the pointer had a built-in level to ensure its two laser lines were exactly horizontal and vertical respectively) and registered the coordinates for subsequent precision analysis (see Fig. 7). At the end of a subject’s experiment we had a session to acquire subjective feedback from the subject. First, there were five questions to be answered and then the subject could comment freely regarding the use of the AR tool and the experiment. Those five questions were:

1. How do you feel to see yourself when you are performing tasks?
2. Do you think the AR tool was helpful?
3. Did the virtual grid make your tasks easier?
4. Do you think you marked the targets correctly both with and without the AR tool?
5. Was it easy to learn the AR tool and how quickly did you adapt yourself to it?

Table 2 Target coordinates (x,y) in cm for Group 2

	Target1	Target2	Target3	Target4	Target5	Target6	Target7	Target8	Target9	Target10
Trial 1	(5,40)	(67,45)	(19,19)	(35,3)	(57,22)	(90,16)	(85,21)	(46,12)	(3,10)	(28,32)
Trial 2	(50,30)	(8,13)	(74,28)	(73,36)	(3,43)	(84,11)	(56,33)	(26,58)	(8,30)	(10,53)
Trial 3	(55,18)	(97,57)	(53,12)	(2,2)	(23,43)	(31,57)	(52,42)	(46,47)	(35,6)	(13,17)



Fig. 7 An experimenter is retrieving the coordinate of a point on the workarea using a ruler and a cross laser pointer

5 Results

5.1 Quantitative measures

Observations in our experiment comprised the target positions marked by subjects and the time they took to mark those targets. The Euclidean distance between the marked positions and their corresponding pre-determined positions was computed as error measure of marking precision, which, together with the aforementioned marking time, formed the quantitative aspect of the experimental results in this study. Coordinates of five marked positions in Group 1 were lost by accident (1 in the manual trial, 4 in the AR without grid trial); for those targets, the error as well as the recorded times were excluded from further analysis. Boxplots in Fig. 8 summarize the distributions of time and error observations of all subjects for the three trials in each group. For clarity, a few extreme outliers were omitted so that plots of the two groups scaled up to the same value range. However, those values were still included in the subsequent statistical analysis. A histogram analysis of the data and Shapiro-Wilk tests confirmed our initial assumption that the radial distance error as well as time observations were not normally distributed. In consequence, we used two-sample Wilcoxon rank-sum tests to ascertain the statistical significance of the differences between these three approaches and the resulting p-values are shown in Tables 3 and 4. More descriptive statistics for both groups are summarized in Tables 5 and 6. Additionally, further analysis of errors did not reveal any salient pattern in terms of error versus target position within the workarea.

5.2 Qualitative results

This section summarizes the answers to the five questions we obtained from the subjects. Although most people gave direct answers to the questions, some did not. Since we did not want to influence their thoughts by forcing answers out of them, we only asked the questions and recorded what they said. Hence, for some questions there do not exist answers from all subjects in the respective group.

Responses of the first group of twelve subjects (university experiment) were as follows: for Question 1, a majority of eight people replied they did not mind seeing themselves in the video during the experiment. Two people said it was strange. One said “it was troublesome

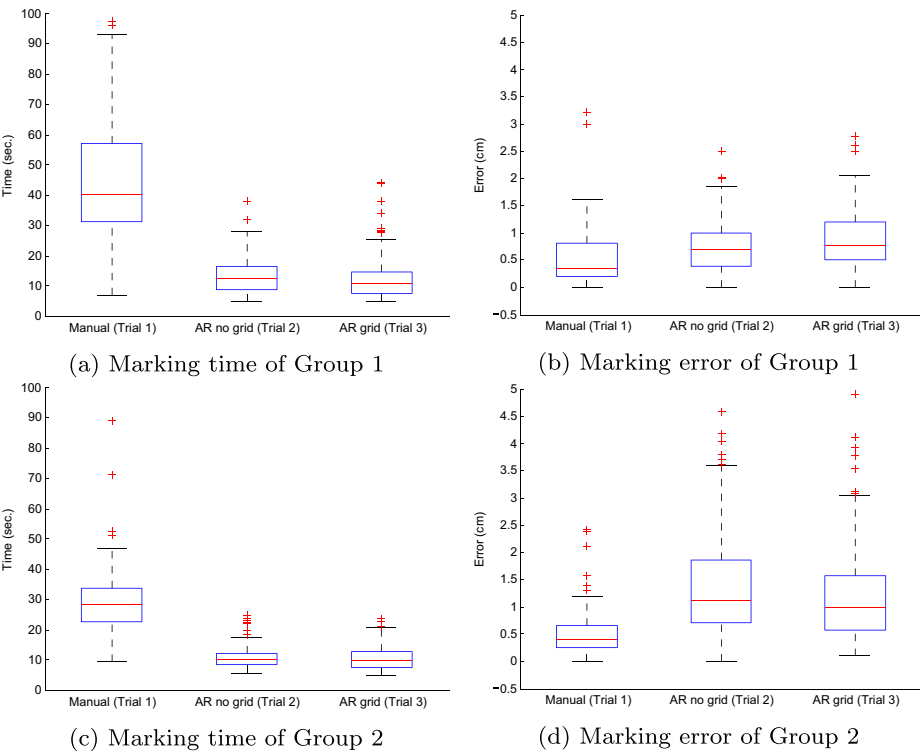


Fig. 8 Marking times and errors of both groups

when I was in the way otherwise no problem.” For Question 2, six people agreed that the AR tool was helpful. Two people only replied that the tool was very fast. One stated it was not reliable, while one was not sure about the tool. For Question 3, the answer was unanimously “no” for all subjects. In regard to Question 4, seven people were less confident with the AR tool while two believed they marked the positions more correctly with it. One subject was not sure about either means and one thought both means offered equal precision. Finally, regarding Question 5, only one person did not think the AR tool was easy to learn.

Group 2 (company experiment) comprised ten people. In response to Question 1, six people said they did not mind seeing themselves in the video. One subject believed it was odd while one only commented that the AR approach was amusing. For Question 2, six people agreed that the tool was helpful while two stated that it would depend on the precision. Six people did not think the grid was helpful while four thought it was (Question 3). In regard to Question 4, six subjects found less confidence in the AR tool while three believed the AR tool was more accurate. Only one thought both means produced the same precision. Finally,

Table 3 The p-values of Wilcoxon rank sum test ($\alpha = 0.05$) for Group 1

	Trial 1 vs. Trial 2	Trial 1 vs. Trial 3	Trial 2 vs. Trial 3
Time	< 0.0001	< 0.0001	0.0797
Error	0.0001944	< 0.0001	0.0345

Table 4 The p-values of Wilcoxon rank sum test ($\alpha = 0.05$) for Group 2

	Trial 1 vs. Trial 2	Trial 1 vs. Trial 3	Trial 2 vs. Trial 3
Time	< 0.0001	< 0.0001	0.4151
Error	< 0.0001	< 0.0001	0.3487

eight people agreed that it was easy to learn this AR tool while one replied that it was easy only for a few points and one believed it was a little hard to learn it (Question 5).

6 Discussions and conclusions

6.1 Marking errors

The Euclidean distance measure (error) is a key performance measure in our study. As the results in Fig. 8b and d as well as Tables 5 and 6 show, users were more precise (lower errors) when using the conventional ruler for target designation and marking compared to the two trials where the AR tool was used. These differences are highly significant, as the very low p-values (Tables 3 and 4) indicate and they were consistently observed in both experiments. We identified two major error sources in the AR trials which contribute to this result. The first error stems from the misalignment between the tip of the pen and the virtual red cross. Certainly a similar misalignment exists between the intended tick on a ruler and the actual place where a subject drew the mark in the manual trial but the small screen of a smartphone with limited resolution undoubtedly aggravates the misalignment. This is in line with the fact that almost half of the subjects from the two groups explicitly commented it was difficult to see the tip of the pen in the AR interface or suggested a zoom-in function. Quite obviously, the users' capability of precisely determining pen positions through the TPP AR interface is limited by the size of the smallest comprehensible image detail on the handheld device. This in turn is affected by several factors such as the pixel resolution and distance of the camera, internal image buffer size for rendering and 3D registration, the resolution of the display of the handheld device etc. With the rapid development of hardware components (both CCD chips and displays), we regard image degradation resulting from current software limitations as the main problematic factor. For example, in our test system, the camera image had to be downsampled to a quarter of its original size, namely 640×360 pixels, for efficient tracking, rendering and transmission through the wireless network. In our experiment, pixel footprint was about three by three millimeters across the workarea considering our camera and system setup. Further reduction of the pixel footprint can be expected given the advances in hardware, which paves way to improved user performance. Regardless of the camera resolution, non-constant pixel footprint sizes may occur

Table 5 Group 1 medians of time, error and 95 % quantile of error

	Trial 1	Trial 2	Trial 3
M. Time (sec.)	40.08	12.31	10.78
M. Error (cm)	0.36	0.69	0.78
Q. Error (cm)	1.34	1.98	1.89

Table 6 Group 2 medians of time, error and 95 % quantile of error

	Trial 1	Trial 2	Trial 3
M. Time (sec.)	28.32	10.10	9.55
M. Error (cm)	0.41	1.11	1.00
Q. Error (cm)	1.40	3.63	3.13

due to perspective distortion when the camera is facing the workarea at a sharp angle, which possibly will give rise to varying user precision across the workarea. Therefore, we recommend employing a camera setup with close to perpendicular alignment when TPP AR is used. A potentially interesting technique of increasing users’ precision might be switching perspectives, e.g. from TPP to FPP in the sense discussed by Salamin et al. [26]), but this would require integration of more complex techniques for both close-range and long-range 3D tracking.

The second source of error, which is inherent to the AR settings, results from transformation. On one hand, there exists mental transformation between the exocentric camera coordinate system presented on display and the egocentric interaction volume of the user. On the other hand, such errors can also manifest themselves through mis-registration caused by the marker-based tracking system (Vuforia in this study), which is an intrinsic part of our AR tool. This aspect of the error can be demonstrated by an extreme case which sporadically occurs when the active marker is partially occluded by the system user (Fig. 9). As shown in the figure, the active marker is on the top left corner and is partly blocked by the user’s head. Since the calibration information is misread by Vuforia, the rendering result of the virtual scene is thus distorted. This may be the cause of some outliers in Trial 2 and Trial 3 and can be difficult to detect when the virtual grid is not rendered, i.e., in Trial 2.

Despite of all potential sources of errors in AR-based target designation and marking, if we compare the median values in Table 5, the differences are merely 0.33 cm and 0.42 cm between Trial 1 and the two AR-based trials respectively. From a more practical perspective, 95 % of the measures from either AR-based trial have errors less than 2 cm according to Table 5. The repeated experiment on Group 2 exhibits similarly small median differences (0.70 cm and 0.59 cm) together with roughly 3 cm error at 95 % quantile according to

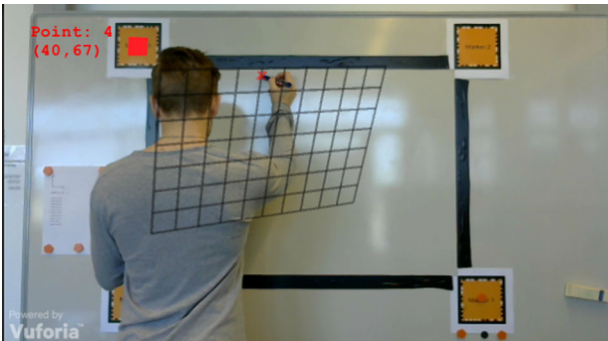


Fig. 9 An example of large tracking error

Table 6. Based on these results, we conclude that TPP AR-based tools are overall viable for designating hidden construction elements in a large number of FM maintenance tasks.

6.2 Efficiency

In Fig. 8a it is obvious that the median times for completing the AR-based trials (Trial 2 and 3) were much shorter than the one for Trial 1. On average, Group 1 performed Trial 2 or Trial 3 at least three times as fast as Trial 1 according to Table 5. These differences are highly significant as well, which is indicated by the small p-values in Table 3. Again, the same pattern and significances were observed in the results of the repeated experiment on Group 2. Although the timing for Trial 1 was done manually, which was different from the other two trials, this did not introduce considerable delay assuming the reaction time for timekeeping is about a second or less as frequently reported. We therefore assert that AR-based target designation is more efficient than a manual approach. This was perceived so as well by a majority of subjects in Group 1, with 8 out of 12 who directly mentioned the AR approach was fast. For a single target, time gains might arguably not be important but we believe that for larger number of targets in practice, the time advantage will pay off. Finally, it should be noted that we did not include the setup time of the AR system in this study, whose largest portion consists of the time required to prepare the real workarea with markers as well as positioning the virtual objects in the scene accordingly. Taking into account the setup time, the efficiency advantage of the current AR tool may not be very obvious comparing with the direct manual measurement. Nevertheless, if an FM task needs to be performed repeatedly on a fixed set of real construction elements over a period of time and with a flexible, well-designed AR system, we believe AR is still a superior solution worth considering.

6.3 Attention and learning

During the tests, experimenters reported that they thought users paid less attention to the task as they were longer into a trial with the AR tool. This observation might agree with the generally much shorter times observed in Trial 2 and Trial 3, as well as with some of the comments from the subjects who stated the tool to be very fast. A block-wise data analysis within trials, however, did not show any strong relation (positive or negative) between errors and stimuli order or time and stimuli order. The absence of such trends suggests that no learning took place in any of the trials. We conclude therefore that the AR-based designation tool is intuitive to use, just as manual marking with a ruler, without the need for much learning. This also agrees with subjects' answers to Question 5, where 19 out of 22 subjects in total felt that the tool was easy to learn.

6.4 User acceptance

From the answers to Question 2 we can conclude that in general people were positive towards this new approach. Although a majority of subjects were less confident with the precision of the AR tool, they still tended to think the tool was helpful. We think the speed of completing the task and the ease of using the tool play an important role here. One interesting point is subjects' reaction towards the virtual grid. Contrary to what we expected, a large portion of subjects did not resort to counting the lines on the grid to locate the target

position but rather dismissed it as “in the way”. We still think that the grid would be a useful function to avoid obviously incorrect measurements due to erroneous AR tracking, because a heavily distorted grid clearly signals to the user that something is wrong.

6.5 Limitations

The current implementation of our AR system did not include any mechanism to handle occlusion between the real and the virtual objects. Since virtual objects are always rendered on top of the video stream, the composite imagery undoubtedly looks unnatural. Correctly resolving the order of the real and the virtual objects as needed for natural occlusion rendering is still a challenging task in the field of AR [4]. Additional sensors such as time-of-flight (ToF) cameras are often employed to provide depth information of the scene, which is unavailable in our experiments. However, since the virtual target (red cross) only occupies a very small region of the screen, its occlusion does not significantly hinder the positioning task.

Another potential limitation for TPP AR is that it is best suited for a reasonably large space to interact with. In the event of narrow space, it is likely that the objects to be augmented do not cover a big area. In such case, the more common FPP AR will be a better alternative for both augmentation and direct interaction. Regarding TPP AR in a larger working space, the area of augmentation and interaction can be easily scaled up by only moving the camera farther away from the area. The downside is that visual quality of tracked features (e.g. markers) degrades with increasing camera distance thus leading to less stable tracking. To counter this, one can increase the size of the tracked features if they are artificial and/or adopt a camera with better resolution. Additionally, considering using a camera with wider field-of-view will also scale up the dimension of area to be augmented without moving the camera away. Finally, in case of very large areas, they can always be subdivided into smaller regions for augmentation.

6.6 Concluding remarks and future work

The analysis of the quantitative results of user experiments have shown that the TPP AR approach towards target designation tasks is at least three times as fast as the manual approach. While the precision of the AR approach is inferior to its manual counterpart, observed errors are still within an acceptable range for a number of AEC/FM tasks such as locating larger artifacts of interest e.g. insulation defects or failure of heating pipes as indicated by infrared thermography, or acquiring the positions of other structural elements like beams and columns. Our findings are in accord with the conclusions drawn in [12], where the AR tool was compared with the traditional total station for steel column inspection tasks. Moreover, subject feedback suggests the TPP does not impose discomfort on users and AR tools employing this perspective seem to be intuitive and easy to use, which are also supported by results from [22]. Future steps in the development of a real system will be towards further reducing sources of error in the AR tool by, among others, adopting a handheld device with a larger display, using video frames with higher resolution and implementing the zoom-in function sought by many subjects. There is also potential to test alternative, more sophisticated marker-less tracking and with such an improved system, finally, to carry out experiments in real scenarios with ground truth targets such as pipes, wires and other construction elements.

Acknowledgments The authors would like to express their gratitude to Anders Blomkvist and Mattias Didriksson, students of IT engineering at Högskolan i Gävle, for system implementation and collection of raw data, and Torsten Jonsson for his practical support and helpful discussions throughout the experiments.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Ahn J, Han R (2012) An indoor augmented-reality evacuation system for the smartphone using personalized pedometry. *Human-Centric Computing and Information Sciences* 2(1):1–23
2. Azuma RT (1997) A survey of augmented reality. *Presence* 6(4):355–385
3. Bae H, Golparvar-Fard M, White J (2013) High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (aec/fm) applications. *Visualization in Engineering* 1(1):1–13
4. Behzadan AH, Dong S, Kamat VR (2015) Augmented reality visualization: a review of civil infrastructure system applications. *Adv Eng Inform* 29(2):252–267
5. Bell B, Höllerer T, Feiner S (2002) An annotated situation-awareness aid for augmented reality. In: *Proceedings of the 15th annual ACM symposium on user interface software and technology*. ACM, pp 213–216
6. Chang KE, Chang CT, Hou HT, Sung YT, Chao HL, Lee CM (2014) Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Comput Educ* 71:185–197
7. Chi HL, Kang SC, Wang X (2013) Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Autom Constr* 33:116–122
8. Côté S, Trudel P (2013) Third person perspective augmented reality for high accuracy applications. In: *IEEE international symposium on mixed and augmented reality (ISMAR)*, 2013. IEEE, pp 247–248
9. Damala A, Cubaud P, Bationo A, Houlier P, Marchal I (2008) Bridging the gap between the digital and the physical: design and evaluation of a mobile augmented reality guide for the museum visit. In: *Proceedings of the 3rd international conference on digital interactive media in entertainment and arts*. ACM, pp 120–127
10. Daponte P, De Vito L, Picariello F, Riccio M (2014) State of the art and future developments of the augmented reality for measurement applications. *Measurement* 57:53–70
11. Dünser A, Billinghurst M, Wen J, Lehtinen V, Nurminen A (2012) Exploring the use of handheld ar for outdoor navigation. *Comput Graph* 36(8):1084–1095
12. Dunston PS, Shin D (2009) Evaluation of augmented reality in steel column inspection. *Autom Constr* 18(2):118–129
13. Dunston PS, Shin DH (2008) Identification of application areas for augmented reality in industrial construction based on technology suitability. *Autom Constr* 17(7):882–894
14. Irizarry J, Gheisari M, Williams G, Walker BN (2013) Infospot: a mobile augmented reality method for accessing building information through a situation awareness approach. *Autom Constr* 33:11–23
15. Kamat VR, El-Tawil S (2007) Evaluation of augmented reality for rapid assessment of earthquake-induced building damage. *J Comput Civ Eng* 21(5):303–310
16. Keil J, Zollner M, Becker M, Wientapper F, Engelke T, Wuest H (2011) The house of olbrich—an augmented reality tour through architectural history. In: *IEEE international symposium on mixed and augmented reality-arts, media, and humanities (ISMAR-AMH)*, 2011. IEEE, pp 15–18
17. Kurkovsky S, Koshy R, Novak V, Szul P (2012) Current issues in handheld augmented reality. In: *International conference on communications and information technology (ICCIT)*, 2012. IEEE, pp 68–72
18. Kwon OS, Park CS, Lim CR (2014) A defect management system for reinforced concrete work utilizing bim, image-matching and augmented reality. *Autom Constr* 46:74–81
19. Milgram P, Colquhoun H (1999) A taxonomy of real and virtual world display integration. *Mixed reality: Merging real and virtual worlds*:5–30
20. Mulloni A, Dünser A, Schmalstieg D (2010) Zooming interfaces for augmented reality browsers. In: *Proceedings of the 12th international conference on human computer interaction with mobile devices and services*. ACM, pp 161–170

21. Mulloni A, Seichter H, Schmalstieg D (2011) Handheld augmented reality indoor navigation with activity-based instructions. In: Proceedings of the 13th international conference on human computer interaction with mobile devices and services. ACM, pp 211–220
22. Nakamura R, Lago LL, Carneiro AB, Cunha AJ, Ortega FJ, Bernardes Jr JL, Tori R (2010) 3pi experiment: immersion in third-person view. In: Proceedings of the 5th ACM SIGGRAPH symposium on video games. ACM, pp 43–48
23. Olbrich M, Graf H, Kahn S, Engelke T, Keil J, Riess P, Webel S, Bockholt U, Picinbono G (2013) Augmented reality supporting user-centric building information management. *Vis Comput* 29(10):1093–1105
24. Rankohi S, Waugh L (2013) Review and analysis of augmented reality literature for construction industry. *Visualization in Engineering* 1(1):1–18
25. Rehl K, Häusler E, Leitinger S, Bell D (2014) Pedestrian navigation with augmented reality, voice and digital map: final results from an in situ field study assessing performance and user experience. *Journal of Location Based Services* 8(2):75–96
26. Salamin P, Thalmann D, Vexo F (2006) The benefits of third-person perspective in virtual and augmented reality? In: Proceedings of the ACM symposium on virtual reality software and technology. ACM, pp 27–30
27. Schall G, Zollmann S, Reitmayr G (2013) Smart vidente: advances in mobile augmented reality for interactive visualization of underground infrastructure. *Pers Ubiquit Comput* 17(7):1533–1549
28. Sukan M, Feiner S, Tversky B, Energin S (2012) Quick viewpoint switching for manipulating virtual objects in hand-held augmented reality using stored snapshots. In: IEEE international symposium on mixed and augmented reality (ISMAR), 2012. IEEE, pp 217–226
29. Tatzgern M, Grasset R, Kalkofen D, Schmalstieg D (2014) Transitional augmented reality navigation for live captured scenes. In: Virtual reality (VR), 2014 IEEE. IEEE, pp 21–26
30. Tatzgern M, Grasset R, Veas E, Kalkofen D, Seichter H, Schmalstieg D (2015) Exploring real world points of interest: design and evaluation of object-centric exploration techniques for augmented reality. *Pervasive Mob Comput* 18:55–70
31. Tönnis M, Plecher DA, Klinker G (2013) Representing information—classifying the augmented reality presentation space. *Comput Graph* 37(8):997–1011
32. Van Krevelen D, Poelman R (2010) A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality* 9(2):1
33. Veas E, Mulloni A, Kruijff E, Regenbrecht H, Schmalstieg D (2010) Techniques for view transition in multi-camera outdoor environments. In: Proceedings of graphics interface 2010. Canadian Information Processing Society, pp 193–200
34. Wagner D, Schmalstieg D (2003) First steps towards handheld augmented reality. In: Proceedings of the 7th international conference on wearable computers (ISWC 2003), pp 127–135
35. White J, Schmidt DC, Golparvar-Fard M (2014) Applications of augmented reality. *Proc IEEE* 102(2):120–123



Fei Liu received his M.Sc. degree in computer science from Department of Information Technology, Uppsala University, Sweden in 2010. He is now a Ph.D. candidate in University of Gävle. His current research interests are mobile visualization and augmented reality.



Stefan Seipel graduated with an MSc in Medical Informatics from Heidelberg University, Germany, in 1992. He obtained a PhD in theoretical medicine from the medical faculty at Heidelberg University, Germany, in 1997 for his thesis on 3D simulation of surgical procedures and surgical instrument navigation. He was appointed professor of Computer Graphics at Uppsala University, Sweden, in 2003.

He currently holds the position of a Professor of Computer Graphics and Head of the research group in Geospatial Information Technology at the University in Gävle, Sweden, as well as he has a part time position as Professor of Computer Graphics at Uppsala University, Sweden. Previous academic posts include Postdoctoral Researcher in Human-Computer Interaction, Senior Lecturer in Graphics, and Senior Lecturer in Visualization. His published research has been directed towards applied computer graphics and interactive visualizations within the medical field and various other fields of applications. His particular research interests include usability of interactive visualizations, spatial and temporal visualization techniques, virtual reality and augmented reality, as well as geographical visualization.